USAID | SENEGAL
DU PEUPLE AMERICAIN

USAID/Senegal All Children Reading (ACR)

Lecture Pour Tous

EGRA CI Analysis Report

Senegal All Children Reading/Lecture Pour Tous
Early Grade Reading Assessment
***Cours d'Initiation*/First Grade Study 2018**

September 5, 2018

Contract no. AID-OAA-I-14-00055/AID-685-TO-16-00003

Produced by: EdIntersect for Lecture Pour Tous
Mr. Pierre de Galbert, Dr. Michel Rousseau, Dr. MaryFaith Mount-Cors

Table of Contents

ACRONYMS AND ABBREVIATIONS

AAM	Assessor Accuracy Measure
ADSE	*Approche Déconcentrée du Suivi Educatif*
AMELP	Activity Monitoring, Evaluation, and Learning Plan
ARED	Associates in Research for Education and Development
CI	*Cours d'Initiation* (First Grade)
CP	*Cours Préparatoire* (Second Grade)
DALN	*Direction de l'Alphabétisation et des Langues Nationales*
DEE	*Direction de l'Enseignement Elémentaire*
DFC	*Direction de la Formation et de la Communication*
EGRA	Early Grade Reading Assessment
IA	*Inspection d'Académie*
IEF	*Inspection de l'Education et de la Formation*
INEADE	*Institut National d'Etude et d'Action pour le Développement de l'Education*
LEMA	Local Education Monitoring Approach
MEL	Monitoring, Evaluation, and Learning
MEN	*Ministère de l'Education Nationale*
MERL	Monitoring, Evaluation, Research, and Learning
PAQUET	*Programme d'Amélioration de la Qualité de de l'Equité et de la Transparence*
SSME	Snapshot of School Management Effectiveness
USAID	United States Agency for International Development

Executive Summary

Lecture Pour Tous is a national reading program led by the Senegalese Ministry of Education (MEN) and supported by USAID. This program supports reading development of students in the first two years of elementary school (CI and CP) in three national languages: Wolof, Pulaar and Seereer. This report presents the results of an Early Grade Reading Assessment (EGRA) conducted with first-grade students as part of the program's Activity Monitoring, Evaluation, and Learning Plan (AMELP). The main objectives of this study were to: (i) provide a measure of basic reading skills of students in Grade 1 in Senegal in three national languages; (ii) provide a measure of Indicator 4 of the AMELP; and (iii) identify the characteristics most closely associated with student reading performance.

Data collection took place in May 2018 in 212 schools across the three language groups in four regions of Senegal: Fatick, Kaffrine, Kaolack, and Matam. The final analytic sample in this study includes 2,166 first-grade students, half of which were girls and half of which were boys, each assessed in the national language chosen for use in the national language reading program in the first two grades at that school. The subtasks included in the assessment in this study were listening comprehension, letter sounds, familiar words, connected-text oral reading, and reading comprehension. Table i below presents the average score and accuracy percentages for each subtask across the three language groups

Table i. Student accuracy and mean scores across language groups

	Wolof		Pulaar		Seereer	
Subtask	Mean Score	Accuracy %	Mean Score	Accuracy %	Mean Score	Accuracy %
Listening Comprehension/% correct answer	71.5		76.6		67.2	
Letter Sounds/ Correct letters per minute	26.4	65.0	24.1	62.7	20.1	57.5
Familiar Words/Correct words per minute	6.2	28.6	3.9	19.6	4.3	21.1

Connected Text Passage/Correct words per minute	5.9	28.2	1.9	9.5	4.9	28.6
Reading Comprehension	14.6	16.8	6.2	7.5	16.8	23.5

When considering the EGRA CI results in relation to the preliminary benchmarks[1] currently in place for Lecture Pour Tous, the overall results show that:

- The majority of students across language groups demonstrate competent **listening comprehension skills**. The average scores in the three language groups range from 67% to 77%, which meets the minimum benchmark for this task. This average range demonstrates competent to high-performing listening comprehension skills and thus a strong foundation of receptive vocabulary in the national languages being used for the national language reading program.
- **Letter sound identification** scores range from 20 to 26 correct letters per minute, placing first-grade students at a beginner stage and, for some, entering a developing stage for understanding the alphabetic principle.
- The average number of **familiar words** read, ranging from 4 to 6 words per minute, indicates first graders are at a beginning to just entering a developing stage in word reading.
- On the **connected text** passage, students in the Pulaar group read an average of 2 words per minute (beginning stage), while their peers in the Seereer and Wolof groups read nearly 5 and 6 words per minute (beginning to developing stages).
- Similarly, students in the Pulaar group correctly answered an average of 6.2% of **reading comprehension questions**, compared with 17% and 15% in the Seereer and Wolof groups respectively. First graders are spanning the spectrum of the beginning stages for reading comprehension.

The main performance indicator of interest for this study is Indicator 4 from the Lecture Pour Tous AMELP: "Average student score for accuracy in reading connected text, measured in the percentage of correct words read of connected grade-level text, in a language students speak and understand (Wolof, Sereer, Pulaar) and at the end of one school year of reading instruction supported by the program." Table ii below presents the overall reading accuracy of students, as well as averages within language groups, gender, and region. These results show that students in the Wolof and Seereer groups were able to read 25% and 22% of words in the passage correctly,

[1] See Table A9 in the annexes to view the preliminary benchmarks.

with the Pulaar group score at 8%. Across language groups, students in Fatick and Kaolack generally read words in the passage more accurately than their peers in Kaffrine and Matam. There were no statistically significant differences between boys and girls on this indicator.

Table ii. Indicator 4 accuracy scores reported overall, by region and gender

	Overall Mean[2]	Wolof	Pulaar	Seereer
Overall	20.5%	24.8%	8.3%*	21.8%
Fatick	24.9%	34.1%	21.8%	19.9%
Kaffrine	17.7%	18.1 %	11.5%	17.6%
Kaolack	25.6%	25.8%	15.5%	27.6%
Matam	7.2%	8.9%*	7.2%*	n/a
Boys	20.0%	25.6%	10.5%	16.6%
Girls	20.9%	23.9%	6.5%	26.9%

* indicates statistical significant differences at the 0.05 level

Given the differences in study design and instruments between this study and the baseline conducted in 2017, there can be no direct comparisons quantifying differences in student performances. It is helpful, however, to contextualize the results of this study by comparing the proportion of students unable to identify one letter or word across subtasks. Table iii below presents the proportion of CI students at baseline and those in this study unable to identify one word or letter correctly. These results show that while 20% to 40% of students were unable to identify a single letter sound in 2017, nearly all students this year could identify at least one letter. In addition, there were almost no CI students able to identify any word at baseline, and this proportion is approximately 50% in the familiar word subtask in this study, and ranges from 40% to 80% in the connected text passage.

Table iii. Proportion of CI students unable to identify one word or letter correctly

	Wolof		Pulaar		Seereer	
	2017	2018	2017	2018	2017	2018

[2] The overall mean across language groups should be interpreted with caution given students were assessed in three different languages depending on the school using different text passages. This indicator is computed to respond to the need of the AMELP to include a program-wide figure.

Letter Sounds	28.9%	5.3%	18.3%	4.8%	39.5%	5.9%
Familiar Words	83.9%	44.4%	77.3%	49.3%	88.1%	56.0%
Connected Text Passage	79.6%	52.1%	91.6%	81.7%	87.9%	37.9%

To reiterate, findings to highlight from this study include:

- Student performance in **listening comprehension** across the three language groups can be categorized as **competent or high performing**.
- Student performance in **letter sound** identification can be categorized as the **developing** level.
- Performance at the **word level and reading comprehension** remains largely at the **beginning** level.
- Students in the Pulaar language group demonstrate lower scores than their peers on average. However, this comparison should be interpreted with caution because the languages are different, and the instruments are different. This finding should encourage the program to investigate reasons for this difference, including Pulaar language variants and vocabulary choice.
- Students in Matam demonstrate lower scores than their peers in other regions, on average. This difference was evident for students learning both in Pulaar and in Wolof (there are no Seereer schools in Matam). However, interestingly, students in Matam learning in Pulaar demonstrated stronger listening comprehension skills than their peers in other regions.
- There were generally no large differences between boys and girls in reading skills. Of note though in the Seereer language group, girls performed better than boys, on average, on letter sounds, familiar words, and reading comprehension.
- The student characteristics most consistently **associated with higher reading accuracy** were: **having books at home and reporting using a book at school on a daily basis.**
- **Teachers and school directors reported high levels of participation** with Lecture Pour Tous, as well as **high levels of satisfaction with the training and teaching materials.**
- Teachers reported **higher levels of confidence in their ability to read and write in the national languages, compared with baseline** reports.

Senegal All Children Reading/Lecture Pour Tous
Early Grade Reading Assessment
Cours d'Initiation/First Grade Study 2018

EGRA CI Analysis Report

Introduction

The EGRA CI Analysis Report presents results of an Early Grade Reading Assessment (EGRA) study that took place in four regions of Senegal where the first year of program implementation for the national language reading program *Lecture Pour Tous* supported by USAID had begun in schools in November 2017[3]. Data collection took place in schools in May 2018. The goal of this study was to measure the status of reading competencies for first grade (*Cours d'initiation* or CI) students in three national languages targeted by Lecture Pour Tous: Wolof, Pulaar, and Seereer.

This study provides data to measure and monitor performance of the Lecture Pour Tous project more frequently than at midline and endline, and to be able to inform program materials and strategy in a timely manner. In addition, this study serves to report a measure for one the main indicators of the Activity Monitoring, Evaluation, and Learning Plan (AMELP) of the program. The main objectives of this study were to: (i) provide a measure of basic reading skills of students in Grade 1 in Senegal in three national languages; (ii) provide a measure of Indicator 4 of the AMELP; and (iii) identify the characteristics most closely associated with student reading performance.

EGRA Baseline Study and EGRA CI (First Grade) Study

This EGRA First Grade Study, or EGRA CI, was referred to as EGRA Light at the outset as it was suggested to be a less-extensive study than the baseline, midline, and endline studies. The baseline study was conducted in May 2017 before the Lecture Pour Tous program started implementation at the school level. Table 1 below presents a comparison of the two types of EGRA studies that will measure student learning under Lecture Pour Tous. The same research

[3] The USAID Activity to support the national reading program is also called Lecture Pour Tous and is referred to as such in this report.

questions were retained (numbers 1 through 3 below) and a similar sample size for schools was maintained as for the baseline in the four regions where data collection took place. The "light" aspect reflects the choice of a smaller number of subtasks for inclusion among the reading items and a streamlined set of interview questions for students, teachers, and school directors. In addition, the EGRA CI uses an instrument designed to be at the first-grade level (developed in February 2018), which is different from the EGRA baseline, midline and endlinee, which uses an instrument at a general second-grade level (developed in February 2017).

Table 1: Comparison of the two types of EGRA studies under Lecture Pour Tous

Baseline, Midline, an Endline EGRA Evaluations	*First Grade or EGRA CI Evaluations*
Conducted by *Lecture Pour Tous* before the project begins implementation, at the midpoint of the project, and in the final year of the project.	Conducted by *Lecture Pour Tous* in the second year of the project, which is the end of the first year of project implementation in schools, and in the fourth year of the project, which is the end of the third year of project implementation in schools.
Used to compare student competencies at baseline to estimate impact of the program. CI (first grade) and CP (second grade) students are sampled.	Used to describe student learning at the end of one year of instruction. CI (first grade) students are sampled.
Does not use counterfactuals.	Does not use counterfactuals.
Data used to report indicator 1 of the AMELP in Year 1, Year 3, and Year 5 of the program.	Data used to report indicator 4 of the AMELP in Year 2 and Year 4 of the program.

Research Questions

The EGRA CI study was aimed at answering the following research questions:

1. What are the basic reading skills of students in Grade 1 in Senegal in three national languages (Wolof, Pulaar, and Seereer)?
    a. What are the differences in performance between boys and girls?
    b. What are the differences in performance between geographic locations (at the region level)?
2. What are the main skills students struggle with?

3. What student, teacher, and school characteristics are associated with higher levels of reading skills?

Sample

This EGRA CI study serves as a measure to evaluate the performance of Lecture Pour Tous after one year. The assessment took place in 212 primary schools across Fatick, Kaffrine, Kaolack, and Matam regions. The sampling was done in two phases, first choosing a representative number of schools in each of the four regions across three language groups, and then randomly sampling 10 CI students (5 boys and 5 girls) from those present on the day of the school visit. After data cleaning, assessments from 2,166 students were maintained in the final dataset. In addition, a small number of CI students who are blind or have low vision were included in a pilot of a braille version of the assessment[4]. These students were part of a separate pilot and are not included in the analytical sample used in this report. Table 2 below provides a description of the total sample analyzed for this report by language group and gender.

Table 2: Sample description by language and gender

	Male	Female	Total
Wolof	466 (48.7%)	491 (51.3%)	957 (100%)
Pulaar	313 (47.7%)	343 (52.3%)	656 (100%)
Seereer	279 (50.5%)	274 (49.5%)	553 (100%)
Total	1,058 (48.9%)	1,108 (51.2%)	2,166 (100%)

EGRA CI Subtasks

As noted above, the EGRA CI instrument was crafted to be at a Grade 1 level, which makes project and ministry staff better able to understand the extent to which students are able to master the letters and types of words and sentences introduced in the Grade 1 teaching and learning materials developed with the support of Lecture Pour Tous. In addition, the Pulaar listening comprehension task used at baseline had to be modified because it was not appropriate across all regional variations. Table 3 below briefly summarizes the subtasks used in this study and the main differences with the 2017 baseline. Tables A7 and A8 in the annexes present a

[4] See *Lecture Pour Tous EGRA CI Braille Pilot Report 2018* for further information on the pilot experience.

complete comparison of the five subtasks used in both studies. As a result of these changes to the instrument, few direct comparisons can be made between the 2017 baseline and this study. Section 5 presents some comparisons that were able to be made with baseline results.

Table 3: Subtasks used in the 2018 EGRA CI Study

Subtask	Number of items	Differences with 2017 Baseline
Listening Comprehension	5 questions	Pulaar passage and questions are new
Letter Sounds	50 letters/graphemes	Fewer letters and order was changed
Familiar Words	25 words	Fewer words and order was changed
Connected Text Passage	19 words (Wolof) 16 words (Pulaar) 20 words (Seereer)	New passage in each language (fewer sentences, shorter sentences, shorter words)
Reading Comprehension	5 questions	New questions with no inferential question

AMELP Indicator 4

The main performance indicator of interest for this study is Indicator 4 from the Lecture Pour Tous Activity Monitoring, Evaluation, and Learning Plan (AMELP): "Average student score for accuracy in reading connected text, measured in the percentage of correct words read of connected grade-level text, in a language students speak and understand (Wolof, Sereer, Pulaar) and at the end of one school year of reading instruction supported by the program." The Indicator 4 score is computed by dividing the number of words read correctly divided by the total words in the passage[5]. In this study, students were given 1 minute to read the connected text passage. The passages were short, from 16 to 20 words depending on the language. If a student were

[5] This differs from a traditional accuracy score where the number of words read correctly is divided by the number of words attempted. In that case, a student who read 5 words accurately may have attempted 10 words and read 5 of those 10 words correctly, which would have resulted in a score of 50% accuracy on the connected text reading passage. If the text has a total of 20 words, the student would have 25% for her Indicator 4 score. On the other hand, a student who attempted to read all 20 words and read 8 of them correctly would have scored only 40% using the traditional methodology.

spending approximately 3 seconds on each word (the norm with EGRA administration rules, i.e. "the 3-second rule"), the student would complete reading the passage within one minute.

In addition to EGRA subtasks and contextual questions for students, survey instruments were administered to 206 school leaders and 207 CI teachers. These brief questionnaires focused on director and teacher characteristics, including their opinion on the use of multiple languages in the early grades and their experience of the professional development they participated in through Lecture Pour Tous. These surveys included a subset of the questions asked at baseline, which were selected either because of specific interest for the program, or because the answers at baseline had suggested they were contextual factors important to student reading development.

The first year of Lecture Pour Tous-supported training for teachers, school directors, and inspectors took place in September and October 2017 and again in February and March 2018 and most teachers had their lesson guides from the beginning of the school year; student materials arrived from Lecture Pour Tous in schools between November 2017 and March 2018. In addition, a teacher strike began in April 2018 and continued for three months in some areas and four months in other areas, ending in June 2018. Anecdotal reports from the field indicate that this strike did not significantly affect Lecture Pour Tous-supported schools.

The findings in this report focus on answering the research questions presented above. The first section below presents an overall summary of results from students, as well as descriptions of teacher, and school characteristics. The second and third sections describe the instrument development and data collection efforts. The fourth, fifth, and sixth sections present detailed results by language group. Section 7 presents these results in context, compared with Lecture Pour Tous AMELP targets and the preliminary student benchmarks, as well as with the 2017 baseline to the limited extent that this is possible. Section 8 describes limitations and section 9 presents a summary of the findings and recommendations for future data collection efforts.

1. Summary of Findings

Table 4 below presents an overall summary of student reading outcomes. For each language group, the table presents student accuracy levels in each subtask as well as the average student

score. The accuracy score is computed by dividing the number of letters or words read correctly divided by the total words attempted. The mean student score for letter sounds, familiar words, and connected text is the number of correct letters or words read per minute. The listening and reading comprehension mean scores are computed as a percentage of correct responses. For listening comprehension, as all students are asked to answer each question after listening to the passage read aloud to them by the assessor, the accuracy and mean score are the same.

Table 4. Student accuracy and mean scores across language groups

	Wolof		Pulaar		Seereer	
Subtask	Mean Score	Accuracy %	Mean Score	Accuracy %	Mean Score	Accuracy %
Listening Comprehension/% correct answer	71.5		76.6		67.2	
Letter Sounds/ Correct letters per minute	26.4	65.0	24.1	62.7	20.1	57.5
Familiar Words/Correct words per minute	6.2	28.6	3.9	19.6	4.3	21.1
Connected Text Passage/Correct words per minute	5.9	28.2	1.9	9.5	4.9	28.6
Reading Comprehension	14.6	16.8	6.2	7.5	16.8	23.5

When considering the EGRA CI results in relation to the preliminary benchmarks[6] currently in place for Lecture Pour Tous, the overall results show that:

- The majority of students across language groups demonstrate competent **listening comprehension skills**. The average scores in the three language groups range from 67% to 77%, which meets the minimum benchmark for this task. This average range demonstrates competent to high-performing listening comprehension skills and thus a strong foundation of receptive vocabulary in the national languages being used for the national language reading program.

[6] See Table A9 in the annexes to view the preliminary benchmarks.

- **Letter sound identification** scores range from 20 to 26 correct letters per minute, placing first-grade students at a beginner stage and, for some, entering a developing stage for understanding the alphabetic principle.
- The average number of **familiar words** read, ranging from 4 to 6 words per minute, indicates first graders are at a beginning to just entering a developing stage in word reading.
- On the **connected text** passage, students in the Pulaar group read an average of 2 words per minute (beginning stage), while their peers in the Seereer and Wolof groups read nearly 5 and 6 words per minute (beginning to developing stages).
- Similarly, students in the Pulaar group correctly answered an average of 6.2% of **reading comprehension questions**, compared with 17% and 15% in the Seereer and Wolof groups respectively. First graders are spanning the spectrum of the beginning stages for reading comprehension.

Drawing a comparison between these results should not lead one to infer that students in one group are necessarily "better" readers than another as the passages were different and the languages themselves are different with distinct alphabets, morphologies, and syntax, and their own progressions for learning to read. In addition, the contexts surrounding each language differ and affect student experiences leading up to their school exposure to the national reading program. In Pulaar in particular, differences in regional variations of the language may still be an important factor.

Results for Indicator 4: accuracy in reading connected text

Table 5 below presents the mean score for Indicator 4 of the Lecture Pour Tous AMELP in each language group overall as well as by region and gender. Overall, the Grade 1 accuracy score (i.e. mean across languages) for connected text was 20.5%. These results show that students in the Wolof and Seereer groups were able to read 25% and 22% of words in the passage correctly, with the Pulaar group score at 8%. Across language groups, students in Fatick and Kaolack generally read words in the passage more accurately than their peers in Kaffrine and Matam. While boys score slightly higher than girls in the Wolof and Pulaar groups, these differences are

not statistically significant. In contrast, girls score higher on this indicator than boys in Seereer, although the difference is also not statistically significant.

Table 5. Indicator 4 accuracy scores reported overall, by region and gender

	Overall Mean[7]	Wolof	Pulaar	Seereer
Overall	20.5%	24.8%	8.3%*	21.8%
Fatick	24.9%	34.1%	21.8%	19.9%
Kaffrine	17.7%	18.1 %	11.5%	17.6%
Kaolack	25.6%	25.8%	15.5%	27.6%
Matam	7.2%	8.9%*	7.2%*	n/a
Boys	20.0%	25.6%	10.5%	16.6%
Girls	20.9%	23.9%	6.5%	26.9%

* indicates statistical significant differences at the 0.05 level

The target for the end of Year Two of the program (after the first year of school-based implementation) set by Lecture Pour Tous was that on average, students should score at least 10% on Indicator 4. Table 6 below presents the proportion of students in each category, by language group, region, and gender who score at least 10% on Indicator 4. Over two-thirds of students from Fatick who are learning to read in Wolof, and three-quarters of their peers in Kaolack who are learning to read in Seereer meet the Year 2 AMELP target. Meanwhile, only about 16% of girls learning to read in Pulaar score 10% or more on Indicator 4.

Table 6. Proportion of students who scored at least 10% on Indicator 4 accuracy scores

	Wolof	Pulaar	Seereer
Overall	53.6%	17.8%	58.6%
Fatick	67.6%	47.1%	53.2%
Kaffrine	36.5%	21.7%	54.8%
Kaolack	44.5%	30.6%	75.6%

[7] The overall mean across language groups should be interpreted with caution given students were assessed in different languages using different text passages. This indicator is computed to respond to the need of the AMELP to include a program-wide figure.

Matam	22.1%	15.4%	n/a
Boys	48.0%	20.1%	49.8%
Girls	44.7%	15.6%	67.3%

SSME[8] survey results: contextual information, attitudes and language skills

Contextual information was collected from school leaders in 206 schools (188 school directors and 18 deputy directors) and CI teachers in 207 schools. The majority of school leaders (98%) and teachers (71%) were men. Educators, both school leaders and teachers, showed a high level of support toward the language policy promoting the teaching of reading in national languages in the early years. School leaders almost unanimously expressed being in total agreement (68%) or agreement (26%) with the policy, compared with only 5% opposed to the policy. Teachers were similarly in total agreement (60%) or agreement (39%) with the policy. However, as seen in the baseline as well, teachers expressed a higher level of comfort with their literacy skills in French than in national languages (see Table 7 below). Similar to the baseline, teachers reported comfort levels higher than 9 out of 10 in French reading, writing, speaking, and oral comprehension skills. Importantly, on the EGRA CI study, teachers reported a higher level of comfort in reading and writing in national languages than at baseline, suggesting the training received and the experience teaching in national languages have led to higher levels of self-efficacy. Teachers were asked as well in the EGRA CI study about their comfort levels in teaching French and national languages and reported an average comfort of 8 for teaching reading in a national language and a comfort level of 9 for teaching reading in French.

The target for the end of Year Two of the program (after the first year of school-based implementation) set by Lecture Pour Tous was that on average, students should score at least 10% on Indicator 4. The target of 10% was met in all categories.

Table 7: Teacher comfort in French and National Languages on a scale of 1 to 10

	Wolof Group	Pulaar Group	Seereer Group

[8] Snapshot of School Management Effectiveness, which includes survey tools for teachers and school directors, as well as short contextual questionnaires for each student who was administered an EGRA test.

	Wolof	French	Pulaar	French	Seereer	French
Reading	7.6	9.2	7.2	9.3	7.9	9.4
Writing	7.1	9.3	7.6	9.3	7.5	9.6
Speaking	8.7	9.3	7.7	9.2	9.0	9.4
Oral Comprehension	8.6	9.2	7.1	8.7	8.9	9.4
Teaching	8.1	8.9	7.7	9.0	8.2	9.0

School directors and teachers reported having an active and positive relationship with Lecture Pour Tous. Nearly all teachers (99%) reported participating in the training and reported being very satisfied (52%) or satisfied (44%). Similarly, 98% of teachers reported having the teacher's reading guide, with 82% finding it very useful and 15% finding it useful. School leaders also reported participating in the training (95%) and being very satisfied (52%) or satisfied (44%). The majority of school directors also reported serving as coaches for CI teachers (79%), and approximately two-thirds reported taking part in CAP (Cellule d'Animation Pédagogique) professional development sessions run by Lecture Pour Tous. These were internal CAPs run by school directors with support from Lecture Pour Tous.

When looking at the association between student characteristics and EGRA outcomes, differences across language groups are noted. Table 8 below presents the association between Indicator 4 scores and eight student characteristics. Overall, a strong association was not detected in this study between reading skills and speaking the national language outside of school. This is largely due to the fact that there is little variation in these variables as most students report speaking the national language used in the classroom with family and friends, especially in the Pulaar and Seereer language groups. The association between parent literacy and student outcomes is positive in the Wolof group, but nearly zero in the two other groups. The variables most consistently associated with positive reading outcomes are owning and using books on a daily basis. Importantly, from a program perspective, using books daily in school is the only variable positively associated with Indicator 4 in all three groups. Students who used books daily in school scored between 7 and 11 percentage points higher on Indicator 4 than their peers who did not[9].

[9] The question asked: How often do you use your "livre-ecole" for learning to read when you are at school? Answer options were: never, one time per month, one time per week, every day.

Table 8: Relationship between student characteristics and Indicator 4

The association between parent literacy and student outcomes is positive in the Wolof group, but nearly zero in the two other language groups.

Student Characteristic	Association with Indicator 4		
	Wolof	Pulaar	Seereer
Speak Wolof/Pulaar/Seereer at home	-5.3	2.1	-2.9
Speak Wolof/Pulaar/Seereer with friends	4.9	2.8	-0.7
Speak Wolof/Pulaar/Seereer with teacher	11.0*	0.1	7.3
Mother can read	13.1*	2.6	0.1
Father can read	10.2*	0.3	1.7
Books to read at home	22.2*	7.9*	7.4
Use school book everyday	10.5*	6.8*	8.6*
Use home book everyday	15.9*	7.6*	3.9

* indicates statistical significant differences at the 0.05 level

2. Instrument Development

The first-grade EGRA instrument was developed in a workshop with similar participants to the baseline instrument workshop participants (representatives from the DEE, INEADE, DALN, ARED, SIL Lead, IEFs of Kaolack, Fatick, Kaffrine, and Matam, and EdIntersect and Lecture Pour Tous) from January 31 until February 9, 2018. The objective of the workshop was do develop the necessary instruments for the EGRA CI study as well as the LEMA (Local Education Monitoring Approach) pilot study to be administered in 2018 during Year Two of Lecture Pour Tous. The expected outcomes of the workshop were connected text passages and reading comprehension passages for each national language of Lecture Pour Tous: Wolof, Pulaar, and Seereer at a first grade level; a listening comprehension passage and questions in Pulaar; the determination of which student contextual questions, and teacher and school director items to retain for the EGRA CI study; the development of classroom observation items that make up the bulk of the LEMA tool; and the validation of the instruments following the field testing of the instruments.

Figure 1: Participants in discussion at the EGRA CI/LEMA instrument development workshop in February 2018.

The workshop activities included four main parts: presentations and discussion in plenary and small group (by language group); a small pilot for field-testing the newly developed subtasks in Kaolack schools in each of the three national languages; discussions about the field testing experience and modifications to the instrument; and finally validation of the instruments in a session with INEADE on February 9, 2018.

After field testing the tasks among Wolof, Pulaar, and Seereer populations in schools in and around Kaolack, the tasks were adapted with more specificity to a first-grade level by the instrument workshop participants. Some key observations from the field testing included: many students in CI were not able to read the texts brought to them for the field test, despite calibrating the texts to CI; the program had differing start dates across schools depending on when books arrived and when students began attending; students were learning three languages starting in CI: the national language at that school, French, and Arabic; some students were not speakers of the national language being used at that school. Upon return to Dakar, the workshop participants worked with a more specific set of guidance including: for connected text subtasks, keep words to no more than two syllables per word; keep sentences to two or three words, with a maximum sentence length of six words; keep sentences simple (no clauses); avoid unnecessary words in the sentence such as superfluous adjectives and adverbs; avoid words from

language variants; use words that contain the first graphemes seen in the program. Participants worked once again in language groups, coming back together at designated times to share and review their work in plenary.

Figure 2: Participants in discussion at the instrument validation session.

The instrument was validated by INEADE, with the DEE, DALN, ARED, SIL Lead, and EdIntersect and Lecture Pour Tous all participating in the session. The final modifications to develop the EGRA CI instrument can be summarized as the following:

- Listening Comprehension: Same texts and questions used with the exception of Pulaar for which a new text and questions were developed avoiding regional variants of the language for a wider comprehensibility across the regions.
- Letter Sounds: Based on the progression in the program, with 50 graphemes included rather than 100 through Week 15 of the curriculum.
- Familiar words: Based on the progression in the program, including 25 words rather than 50.
- Connected text reading: adjusted to a CI level as described above. Total length no more than 20 words.
- Reading comprehension questions: adjusted to a CI level.

- Student context questions: 11 questions retained, with questions prioritized that have been associated previously with reading performance and/or with program implementation information.
- Teacher questions: 13 questions retained, with questions prioritized that have been associated previously with student reading performance and/or with program implementation information.
- School director questions: 10 questions retained (with two of those being multi-part questions), with questions prioritized that have been associated previously with student reading performance and/or with program implementation information.

A full comparison of the EGRA CI and full EGRA used at baseline is included in tables A7 and A8 in the Annex.

3. Data Collection

As with the baseline study, the methodology consisted of two types of data collection instruments: Early Grade Reading Assessment (EGRA) and Snapshot of School Management and Effectiveness (SSME) questionnaires. The enumerator training was held in Dakar from May 22-16, 2018. A total of 62 trainees attended the training together with Lecture Pour Tous staff and representatives of the MEN. During the training, enumerators were instructed in EGRA, SSME questionnaires, and sampling procedures. For the EGRA subtasks, goals and rules were presented in French, and then practice was conducted in the 3 languages. Each trainer from INEADE was responsible for one language group and plenary sessions were conducted together with language-based sessions. Two Assessor Accuracy Measure (AAM) exercises were conducted during the training to determine the participants' accuracy. A quiz took place on day five of the training in order to assess enumerators' knowledge of the topics covered.

Data collection started on May 29, 2018 and continued for nine days until June 8, 2018. The assessors were divided into 24 teams of 2 or 3, each responsible to collect data in 9 to 11 schools. The data collection exercise aimed to assess 10 students sampled from first-grade classrooms (CI) in 212 schools across the three language groups. Schools were sampled from a population of 1,115 schools covering 4 regions. The sampling strategy took into consideration the goal of assessing reading levels within a language group and comparisons within language group across grades and gender and IEF[10].

[10] See *Lecture Pour Tous EGRA Training and Data Collection Report 2018* for further information about EGRA CI data collection.

4. Findings Specific to Wolof Language Group

Students in the Wolof group presented, on average, competent listening comprehension skills, developing letter sound skills, and beginning word and connected text reading and reading comprehension. Table 9 presents student mean scores, accuracy, and 95% confidence intervals for students in the Wolof group for each subtask. Students answered on average 72% of the listening comprehension questions correctly. Student skills were developing at the letter sound level, with an average of 26 letters per minute correct, but at a beginning stage at the word level at approximately 6 words per minute in both the familiar words and words in a connected text. Accuracy was also much higher with letter sounds, where students can correctly identify more than two thirds of the letters they attempt, on average.

Table 9. Wolof student average scores and accuracy by subtask

Subtask	Average Score		Accuracy	
	Mean	95% CI	Mean	95% CI
Listening Comprehension/ % Correct answer	71.5	68.2 ; 74.7		
Letter Sounds / Correct letters per minute	26.4	23.9 ; 28.8	65.0	60.5 ; 69.5
Familiar Words / Correct words per minute	6.2	5.2 ; 7.3	28.6	24.7; 32.6
Connected Text Passage	5.9	4.9 ; 6.9	28.2	24.3; 32.2
Reading Comprehension / % Correct answer	14.6	11.8 ; 17.5	16.8	13.8; 19.7
Indicator 4	24.8	21.1 ; 28.4		

Tables 10 and 11 below present the average scores by gender and region. Overall, boys and girls in the Wolof group performed at similar levels across the subtasks. There were only small differences in all subtasks, and these were not statistically significant. At the regional level, students in Fatick and Kaolack performed better overall than their peers in Kaffrine and Matam. In particular, students in Matam scored statistically lower than their peers in all three regions for listening comprehension and familiar words, as well as lower than those in Fatick in connected words, reading comprehension, and on Indicator 4. Students in Kaffrine scored lower than their peers in Fatick on reading comprehension questions and Indicator 4.

Table 10. Wolof students' average scores by subtask and gender

Subtask	Male		Female	
	Mean	95% CI	Mean	95% CI
Listening Comprehension/ % Correct answer	71.0	66.7 ; 75.4	71.6	67.7; 75.4
Letter Sounds / Correct letters per minute	25.5	23.2; 27.9	26.7	23.9; 29.6

Familiar Words / Correct words per minute	6.5	5.3; 7.7	6.0	4.8; 7.1
Connected Text Passage	6.1	4.9; 7.2	5.4	3.8; 6.9
Reading Comprehension / % Correct answer	15.6	12.3 ; 18.8	14.0	10.7 ; 17.3
Indicator 4	25.6	21.6 ; 29.6	23.9%	19.3 ; 28.5

* indicates statistical significant differences between male and female at the 0.05 level

Table 11. Wolof students' average scores by subtask and region

	Fatick	Kaffrine	Kaolack	Matam
Subtask	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
Listening Comprehension/ % Correct answer	70.3 (65.3 ; 75.3)	71.9 (67.9 ; 76.0)	71.8 (66.1 ; 77.6)	50.3* (42.7 ; 58.0)
Letter Sounds / Correct letters per minute	33.9 (29.9 ; 38.1)	21.6 (18.3 ; 24.9)	26.2 (22.1 ; 30.2)	15.4 (6.1 ; 24.6)
Familiar Words / Correct words per minute	8.8 (6.6 ; 11.0)	4.3 (3.3 ; 5.3)	6.6 (4.8 ; 8.4)	1.2* (0.0 ; 2.4)
Connected Text Passage	7.6 (5.6 ; 9.6)	4.3* (3.3 ; 5.4)	6.0 (4.0 ; 7.9)	1.9* (0.0 ; 4.1)
Reading Comprehension / % Correct answer	20.6 (13.5 ; 27.7)	8.7* (6.3 ; 11.1)	16.9 (12.0 ; 21.8)	3.9* (0.0 ; 9.3)
Indicator 4	34.1% (26.2 ; 42.0)	18.1%* (13.9 ; 22.2)	25.8% (19.5 ; 32.1)	8.9%* (0.0 ; 18.9)

* indicates statistical significant differences between some regions at the 0.05 level

Table 12 presents the mean proportion of students who answer positively to eight student characteristics from the questionnaire, as well as the association between these characteristics and Indicator 4. The large majority of students in the Wolof group speak Wolof at home, school, and with their teacher. In contrast, few students report their parents are able to read. Finally, students generally report having books at home, as well as using books on a daily basis both at home and at school. The characteristics are all positively associated with Indicator 4, with the exception of speaking Wolof at home and with friends, for which there is no significant difference.

Table 12. Relationship between Wolof student characteristics and Indicator 4

	Proportion	Association with Indicator 4	95% CI
Speak Wolof at home	77%	-5.3	-13.8 ; 3.1

Speak Wolof with friends	88%	4.9	-4.0 ; 13.8
Speak Wolof with teacher	87%	11.0*	3.9 ; 18.1
Mother can read	15%	13.1*	6.6 ; 19.6
Father can read	17%	10.2*	2.6 ; 17.8
Books to read at home	89%	22.2*	17.3 ; 27.2
Use school book everyday	79%	10.5*	3.9 ; 17.1
Use home book everyday	79%	15.9*	9.2 ; 22.7

* indicates statistical significant differences at the 0.05 level

5. Findings Specific to Pulaar Language Group

Similar to their peers learning to read in Wolof, students in the Pulaar group presented, on average, competent listening comprehension skills, developing letter sound skills, and beginning word and connected text reading skills and reading comprehension. Table 13 presents student mean scores, accuracy, and 95% confidence intervals for students in the Pulaar group for each subtask. Students answered on average 77% of the listening comprehension questions correctly. Student skills were developing at the letter sound level, with an average of 24 letters per minute correct, but developing at the word level with fewer than 4 and 2 words per minute in the familiar words and words in a connected text respectively. Accuracy was also much higher with letter sounds, where students could correctly identify nearly two thirds of the letters they attempted, on average, compared with 20% of familiar words and 10% of words in the connected passage.

Table 13. Pulaar student average scores and accuracy by subtask

Subtask	Mean Score		Accuracy	
	Mean	95% CI	Mean	95% CI
Listening Comprehension/ % Correct answer	76.6	72.7 ; 80.4		
Letter sound / Correct letters per minute	24.1	21.2 ; 26.9	62.7	57.2 ; 69.3
Familiar word / Correct words per minute	3.9	2.9 ; 4.8	19.6	15.3 ; 23.9
Connected Text Passage	1.9	1.2 ; 2.6	9.5	0.6 ; 12.8
Reading Comprehension / % Correct answer	6.2	3.8 ; 8.7	7.5	4.6 ; 10.4
Indicator 4	8.3	5.4 ; 11.5		

The new Pulaar listening comprehension subtask was designed to eliminate regional vocabulary and various Pulaar version differences so as to determine if students in the program had listening

comprehension abilities in a simplified and non-version-specific Pulaar. As can be seen above, the new passage is clearly more understandable for the range of Pulaar speakers in Senegal than was the baseline task. Students in this group demonstrated receptive comprehension much more closely related to their peers in other language groups in 2018 than at baseline.

Tables 14 and 15 below display student scores across gender and region. Similar to students in the Wolof group, while there are small differences across gender, there are not statistically significant differences between boys and girls in fluency scores and accuracy. Comparisons across regions are surprising and worth exploring further. Students learning to read in Pulaar in Fatick, Kaffrine and Kaolack score the same across subtasks, on average. Students in Matam, in contrast, score significantly higher than their peers in all three regions on listening comprehension skills, but lower than their peers on all other subtasks. In particular, they score significantly lower than students in Fatick on Indicator 4.

Table 14. Pulaar students' average scores by subtask and gender

Subtask	Male		Female	
	Mean	95% CI	Mean	95% CI
Listening Comprehension/ % Correct answer	76.7	71.9 ; 81.5	77.2	72.7 ; 81.7
Letter Sounds / Correct letters per minute	25.4	22.1 ; 28.8	22.8	19.1 ; 26.5
Familiar Words / Correct words per minute	4.3	2.9 ; 5.6	3.4	2.4 ; 4.5
Connected Text Passage	2.1	0.9 ; 3.3	1.6	0.8 ; 2.3
Reading Comprehension / % Correct answer	7.4	3.2 ; 11.7	5.0	2.6 ; 7.4
Indicator 4	10.5	5.1 ; 15.9	6.5	3.7 ; 9.3

* indicates statistical significant differences between male and female at the 0.05 level

Table 15. Pulaar students' average scores by subtask and region

	Fatick	Kaffrine	Kaolack	Matam
Subtask	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
Listening Comprehension/ % Correct answer	50.9 (40.6 ; 61.2)	65.8 (60.1 ; 71.4)	64.9 (58.0 ; 71.8)	79.5* (75.1 ; 83.9)
Letter Sounds / Correct letters per minute	26.9 (21.2 ; 32.5)	22.5 (16.9 ; 27.9)	23.0 (19.6 ; 26.4)	23.2 (20.8 ; 27.6)
Familiar Words / Correct words per minute	6.2 (3.9 ; 8.5)	6.1 (2.9 ; 9.4)	5.9 (3.9 ; 7.8)	3.5 (2.4 ; 4.5)
Connected Text Passage	4.1	5.7	3.3	1.4

	(2.1 ; 6.0)	(0.0 ; 11.5)	(1.8 ; 4.8)	(0.7 ; 2.2)
Reading Comprehension / % Correct answer	11.5 (4.8 ; 18.0)	7.8 (3.4 ; 12.3)	9.1 (5.6 ; 12.5)	5.7 (2.8 ; 8.6)
Indicator 4	21.8% (13.3 ; 30.4)	11.5% (4.8 ; 18.1)	15.5% (9.4 ; 21.7)	7.2* (3.7 ; 10.7)

* indicates statistical significant differences between some regions at the 0.05 level

Table 16 presents the mean proportion of students who answer positively on a range of student characteristics from the questionnaire, as well as the association between these characteristics and Indicator 4. Approximately nine out of ten students in this group speak Pulaar at home and with their friends, and more than two thirds speak it with their teacher. The lower proportion of students who report speaking Pulaar with their teacher than their peers speaking Wolof with their teacher should be noted and further explored. Very few students report their parents being able to read. Finally, students generally report having books at home, as well as using books on a daily basis both at home and at school. In contrast with the Wolof group, few student characteristics are associated with higher reading accuracy. Having books at home and using books on a daily basis are associated with higher accuracy.

> Having books at home and using books on a daily basis are associated with higher accuracy.

Table 16. Relationship between Pulaar student characteristics and indicator 4

	Proportion	Association with indicator 4	95% CI
Speak Pulaar at home	92%	2.1	-3.5 ; 7.6
Speak Pulaar with friends	89%	2.8	-1.9 ; 7.4
Speak Pulaar with teacher	67%	0.1	-4.2 ; 4,3
Mother can read	8%	2.6	-6.0 ; 11.2
Father can read	6%	0.3	-6.0 ; 6.6
Books to read at home	81%	7.9*	3.7 ; 12.0
Use school book everyday	90%	6.8*	3.5 ; 10.0
Use home book everyday	82%	7.6*	3.4 ; 11.9

* indicates statistical significant differences at the 0.05 level

6. Findings Specific to Seereer Language Group

The performance of students in the Seereer group presents more similarities with their peers in the Wolof group than in the Pulaar group. These students presented, on average, competent to high-performing listening comprehension skills, with an average of two-thirds of the listening comprehension questions answered correctly. Students identified 20 letters correctly per minute, and read 4 to 5 words correctly per minute in the familiar word and connected text reading passage subtasks. Accuracy scores are higher across subtasks, with students correctly identifying 58% of the letters they attempted, on average, compared with 21% of familiar words and 29% of words in the connected text passage. Table 17 presents student mean scores, accuracy, and 95% confidence intervals for students in the Seereer group in each subtask.

Table 17. Seereer students' average scores and accuracy by subtask

Subtask	Mean Score		Accuracy	
	Mean	95% CI	Mean	95% CI
Listening Comprehension/ % Correct answer	67.2	63.1 ; 71.2		
Letter Sounds / Correct letters per minute	20.1	17.6 ; 22.7	57.5	52.2 ; 62.9
Familiar Words / Correct words per minute	4.3	3.3 ; 5.3	21.1	16.5 ; 25.7
Connected Text Passage	4.9	3.8 ; 6.0	28.6	22.9 ; 34.3
Reading Comprehension / % Correct answer	16.8	12.7 ; 20.9	23.5	17.8 ; 29.3
Indicator 4	21.8	17.1 ; 26.4		

Tables 18 and 19 below display student scores across gender and region for the Seereer group. In contrast to the two other language groups, there seems to be a difference in student performance across gender. While there is no significant difference in listening comprehension, girls performed better on average in the four other subtasks. These differences are statistically significant for letter sounds, familiar words, and reading comprehension, but not for the connected text subtask and Indicator 4. At the regional level, students in Kaolack scored higher overall, but there were no significant differences.

Table 18. Seereer students' average scores by subtask and gender

Subtask	Male		Female	
	Mean	95% CI	Mean	95% CI
Listening Comprehension/ % Correct answer	66.3	60.7 ; 71.8	68.0	63.6 ; 72.4

Letter Sounds / Correct letters per minute	17.2	14.8 ; 19.6	23.0*	19.9 ; 26.2
Familiar Words / Correct words per minute	3.1	2.2 ; 3.9	5.5*	4.0 ; 6.9
Connected Text Passage	3.6	2.6 ; 4.6	6.2	4.6 ; 7.7
Reading Comprehension / % Correct answer	11.7	8.2 ; 15.1	21.9*	16.1 ; 27.6
Indicator 4	16.6%	12.2 ; 21.1	26.9%	20.8 ; 32.9

* indicates statistical significant differences between male and female at the 0.05 level

Table 19. Seereer students' average scores by subtask and region

	Fatick	Kaffrine	Kaolack	Matam
Subtask	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
Listening Comprehension/ % Correct answer	68.2 (64.4 ; 71.9)	72.4 (62.7 ; 82.0)	63.6 (52.7 ; 74.6)	-
Letter Sounds / Correct letters per minute	19.7 (16.4 ; 23.0)	15.9 (6.5 ; 25.3)	21.7 (19.3 ; 24.1)	-
Familiar Words / Correct words per minute	3.9 (2.7 ; 5.2)	4.3 (1.0 ; 7.6)	5.4 (4.0 ; 6.8)	-
Connected Text Passage	4.5 (3.1 ; 5.9)	3.5 (1.0 ; 6.0)	6.2 (4.7 ; 7.7)	-
Reading Comprehension / % Correct answer	15.5 (10.7 ; 20.4)	12.2 (2.6 ; 22.0)	20.9 (14.5 ; 27.5)	-
Indicator 4	19.9% (14.1 ; 25.9)	17.6% (5.2 ; 30.1)	27.6% (22.2 ; 32.9)	-

* indicates statistical significant differences between some regions at the 0.05 level

Table 20 presents the mean proportion of students who answer positively to eight student characteristics from the questionnaire, as well as the association between these characteristics and Indicator 4. The vast majority of students speak Seereer at home, with their friends, and with their teacher. A low proportion of students report their parents are able to read. Finally, students generally report having books at home, as well as using books on a daily basis both at home and at school. The only characteristic positively associated with Indicator 4 is the daily use of books in school.

The only characteristic positively associated with Indicator 4 is the daily use of books in school.

Table 20. Relationship between Seereer student characteristics and Indicator 4

	Proportion	Association with indicator 4	95% CI
Speak Seereer at home	92%	-2.9	-22.2 ; 16.4
Speak Seereer with friends	92%	-0.7	-16.2 ; 14.8
Speak Seereer with teacher	88%	7.3	-1.1 ; 15.6
Mother can read	12%	0.1	-8.1 ; 8.3
Father can read	16%	1.7	-5.9 ; 9.3
Books to read at home	94%	7.4	-0.9 ; 15.7
Use school book everyday	89%	8.6*	1.6 ; 15.5
Use home book everyday	86%	3.9	-8.2 ; 16.0

* indicates statistical significant differences at the 0.05 level

7. Contextualizing results

Analysis in relation to the EGRA baseline

As explained above, the results from the EGRA CI 2018 study cannot be directly compared to the 2017 baseline study because the instruments included major modifications, with the exception of listening comprehension in Wolof and Sereer. It is possible, however, to present these results in context by describing the proportion of zero scores in both studies and how the students performed according to the preliminary benchmarks established by Lecture Pour Tous in 2017. These comparisons should be made cautiously, and the magnitude of differences should not be made using figures, but by describing trends broadly.

> Approximately half of students can decode at least one familiar word in all three languages, as well as words in a connected text passage.
>
> A notable exception is that the proportion of students in the Pulaar group unable to read a single word in the connected text passage remains very high, at 82%.

Tables 21, 22, and 23 present a comparison of zero scores between the 2017 and 2018 studies in each language group, with the proportion and the 95% confidence interval. For listening comprehension in Wolof and Seereer, the only two valid direct comparisons as those tasks stayed exactly the same from baseline to the EGRA CI study, the proportion of students unable to answer any questions correctly remains very low, and there are no statistically significant differences. In Pulaar, the

proportion of zero scores has drastically fallen to 2.7%, a proportion similar to that in the two other groups, indicating the new instrument is a better measure of students' receptive comprehension skills.

Across the three language groups, the proportion of students unable to identify a single letter is not very low, with between 5 and 6 percent zero scores. This drop may be explained both by the fact that students at baseline had never been taught in these languages, and that the instrument now focuses specifically on the letters/graphemes included in the CI materials supported by Lecture Pour Tous. Nevertheless, this positive outcome should be noted. Similarly, an important decline in the proportion of students unable to decode a single word should be noted, even while considering that the instruments were simplified. Approximately half of students can decode at least one familiar word in all three languages, as well as words in a connected text passage. A notable exception is that the proportion of students in the Pulaar group unable to read a single word in the connected text passage remains very high, at 82%.

Table 21. Proportion of Wolof students with zero scores by subtask

Subtask	EGRA CI 2018		EGRA BASELINE 2017 (CI students only)	
	%	95% CI	%	95% CI
Listening Comprehension	5.6%	3.9 ; 8.0	4.5%	2.5 ; 7.9
Letter Sounds	5.3%	3.6 ; 7.7	28.9%	23.4 ; 35.0
Familiar Words	44.4%	38.6 ; 50.5	83.9%	79.5 ; 87.6
Connected Text Passage	52.1%	45.9 ; 58.3	79.6%	72.8 ; 85.1
Reading Comprehension	69.5%	64.4 ; 74.1	91.3%	87.4 ; 94.1

Table 22. Proportion of Pulaar students with zero scores by subtask

Subtask	EGRA CI 2018		EGRA BASELINE 2017	
	%	95% CI	%	95% CI
Listening Comprehension	2.7%	1.3 ; 5.5	56.3%	46.1 ; 65.9
Letter Sounds	4.8%	2.3 ; 9.6	18.3%	13.1 ; 24.9
Familiar Words	49.3%	40.3 ; 58.3	77.3%	70.0 ; 83.1
Connected Text Passage	81.7%	75.2 ; 86.7	91.6%	86.7 ; 94.8
Reading Comprehension	85.7%	79.8 ; 90.1	98.6%	95.6 ; 99.5

Table 23. Proportion of Seereer students with zero scores by subtask

Subtask	EGRA CI 2018		EGRA BASELINE 2017	
	%	95% CI	%	95% CI
Listening Comprehension	3.5%	1.9 ; 6.3	8.9%	5.3 ; 14.6
Letter Sounds	5.9%	3.5 ; 9.9	39.5%	33.0 ; 46.4
Familiar Words	56.0%	47.8 ; 63.9	88.1%	83.2 ; 91.6
Connected Text Passage	37.9%	30.5 ; 45.9	87.9%	82.1 ; 92.1
Reading Comprehension	61.3%	53.7 ; 68.5	96.2%	91.5 ; 98.3

Analysis in relation to the preliminary student performance standards and benchmarks

Based on the 2017 baseline data, Lecture Pour Tous helped the ministry to develop a preliminary set of standards and benchmarks to categorize the performance of students across subtasks. Table A9 in the annexes presents the full set of preliminary benchmarks for accuracy, fluency, and comprehension for CI and CP. Using the CI benchmarks, Figures 3, 4 and 5 below show the proportion of students in each category across the three language groups.

Across the three language groups, the distribution of students on the listening comprehension is balanced, with at least half of students categorized as competent or high performing. The distribution of students across categories for letter accuracy is also balanced. The majority of students in all three groups fall in the beginner category for all other subtasks.

Across the three language groups, the distribution of students on the listening comprehension is balanced, with at least half of students categorized as competent or high performing.

The distribution of students across categories for letter accuracy is also balanced.

The majority of students in all three groups fall in the beginner category for all other subtasks.

Figure 3. Performance Categories – Wolof

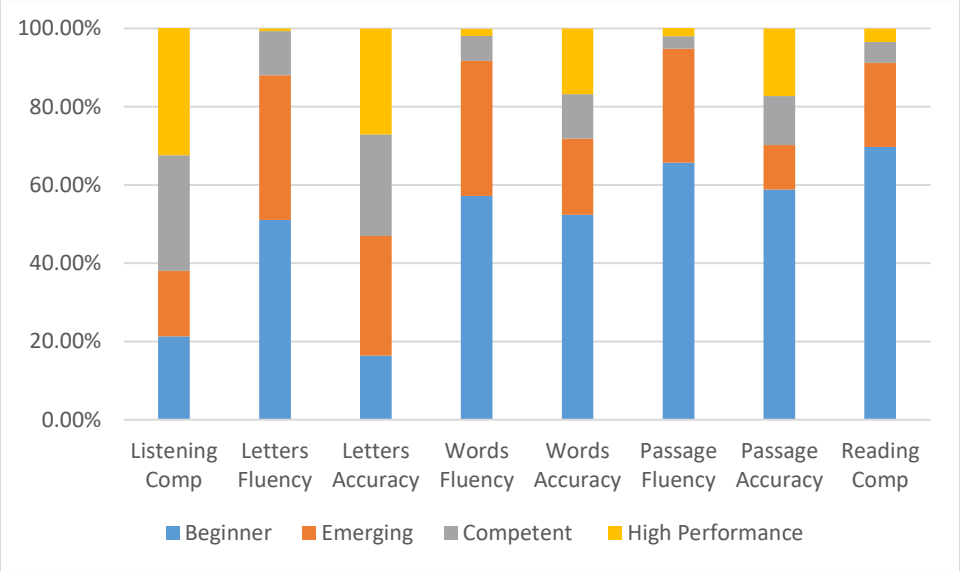


Figure 4. Performance Categories – Pulaar

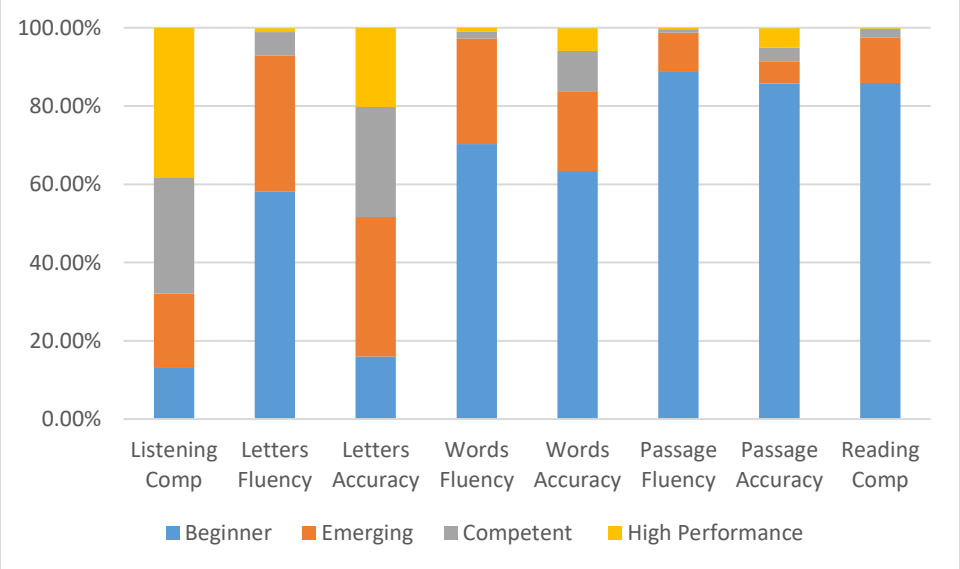
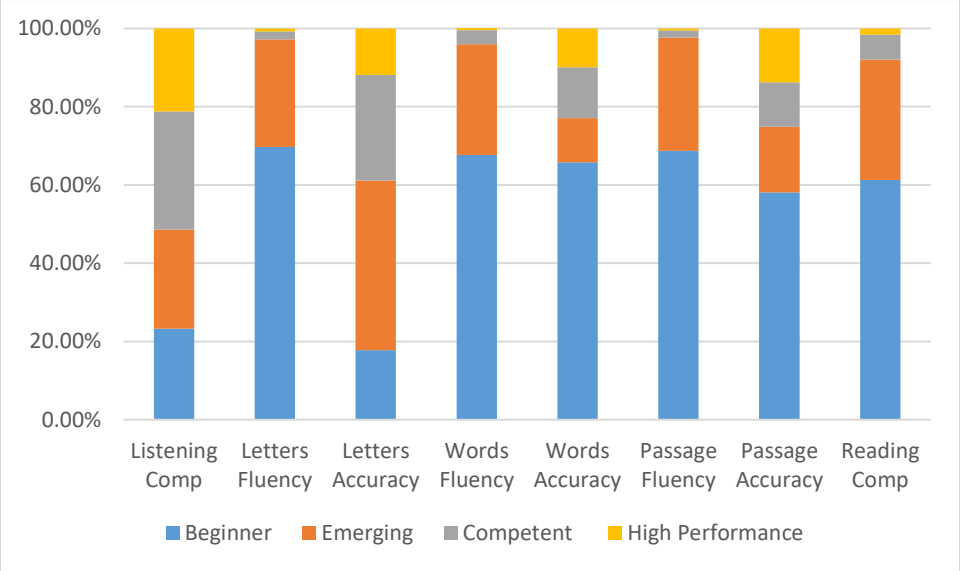


Figure 5. Performance Categories – Seereer



8. Limitations and Considerations

The main objectives of the EGRA CI study were largely met, including the development of new instruments, and data collection in a sample of schools across the three language groups large enough to make inferences about overall proficiency of first grade students, as well as comparisons across gender and regions. Nevertheless, a number of limitations and considerations are important to take into consideration when interpreting the findings of the study. These fall under two categories: data collection challenges, and the measurement and interpretation of Indicator 4 of the AMELP.

Data collection

- **Data collection calendar**: In some regions, planning of data collection was not properly coordinated, resulting in some teams being spread out across the region making supervision difficult due to long distances between schools. For the next rounds of data collection, it would be advisable to have the calendar ready before the start of the enumerator training in order to properly plan the supervision calendar;
- **Duration of data collection in each school**: Conducting data collection in one day proved to be an important improvement in order to increase the quality of the data; this arrangement should be kept for next rounds of data collection, which may lead to

increasing the team size when assessing more students at each school (in this year's EGRA CI study, data collection was only limited to CI);

- **Team size:** Teams of 2 or 3 are preferred over single enumerators going to complete the data collection in one school. Having bigger teams provides better supervision possibilities and reduces the amount of mistakes done while collecting data (thanks to peer-to-peer feedback). For the next rounds of data collection, teams should be kept to a minimum of 2 or 3 members (according to sample size) and in the case of unavailability of enumerators, data collection duration could be extended;
- **Supervision:** In-depth, quality supervision was possible this year due to the outstanding job done by Lecture Pour Tous field agents and field offices. Supervision was managed well, and it was possible to visit teams multiple times over the course of data collection. Though mistakes on the part of some enumerators remain, feedback given improved the quality of the administration.
- **INEADE participation.** It was unfortunately not possible to have the three INEADE trainers take part in the supervision of the data collection due to administrative issues.

Indicator 4

- **Indicator development.** Given the overlapping timelines between the development of the AMELP performance indicators and the instrument development for the EGRA CI study, communication between the program staff and the research partner was not optimal. This resulted in a gap in understanding on how to compute Indicator 4 exactly as it is described in the Performance Indicator Reference Sheet (PIRS).
- **Defining "no time limit".** The main consideration that requires clarification is the definition of a time limit. The PIRS requires students to read a passage with "virtually no time limit" in order to measure student accuracy, and further defines virtually as "up to three minutes." The EGRA CI instruments developed included passages that had a maximum of 20 words. One of the common rules of EGRA administration is "the 3 second rule", encouraging students to move to the next word if they cannot read a word in 3 seconds. This allows students who would otherwise stay stuck on one word to attempt more words. Given the maximum length of the connected texts in this study, students should all have attempted the last word of the text (3 seconds x 20 words = 1 minute)[11]. However, some students did not have the time to attempt all words because of the uneven enforcement of the 3 second rule by assessors. This resulted in a small discrepancy between Indicator 4 and the traditional measure of accuracy presented in

[11] The only cases for which this is not the case is the early stop-rule, when students are unable to read a single word from the first line of the text.

Tables 4, 9, 13, and 17. A decision should be made as to how to best measure Indicator 4 managing to maintain both the notion of "virtually no time limit" and EGRA administration rules. The most practical solution may be to maintain the 3 second rule and allow students to take up to three minutes to read the passage.

9. Summary and Recommendations

The EGRA CI study conducted in May 2018 has shown that students on average at the end of first grade after the first year of Lecture Pour Tous activities at the school level have competent listening comprehension skills, beginning letter sound (alphabetic principle) skills, and beginning word reading skills. With the exception of letter sounds, where students are able to identify upwards of 20 letters per minute, students at the end of first grade are just beginning to read individual words and words connected in simple sentences. As a result, reading comprehension scores are at a beginning level, on average. Despite the limitations in comparing these results directly to the 2017 baseline, it should be noted that with the instrument now designed for the Grade 1 level, Grade 1 students displayed stronger skills in schools that received support from Lecture Pour Tous for one year than Grade 1 students at baseline. Furthermore, Grade 1 students met the Lecture Pour Tous Year 2 program target of reading connected text with at least 10% accuracy. This should be considered encouraging after a first year for the program during which material distribution was delayed, and some coaching from inspectors was affected by the strikes.

Despite the limitations in comparing these results directly to the 2017 baseline, it should be noted that with the instrument now designed for the Grade 1 level, Grade 1 students displayed stronger skills in schools that received support from Lecture Pour Tous for one year than Grade 1 students at baseline. Furthermore, Grade 1 students met the Lecture Pour Tous Year 2 program target of reading connected text with at least 10% accuracy.

From a programmatic perspective, several important points can be noted. First, the feedback from teachers and school directors indicates both high levels of participation and appreciation for the training under Lecture Pour Tous. In addition, when asked about their ability in French and national languages, teachers reported higher levels of confidence than at baseline prior to taking part in the training. Finally, one of the important student characteristics associated with

positive student outcomes on Indicator 4, daily use of books in schools, is a pedagogical practice that can be stressed through professional development.

The midline EGRA data collection planned for May 2019 should carefully consider the sampling of schools and choice of instruments to reflect the goals of the study and the change in implementation rollout. At the end 2018-2019 school year, students in CP will have benefited from Lecture Pour Tous in approximately half of the schools in the four regions of Fatick, Kaolack, Kaffrine, and Matam. CI students and teachers in all schools in these regions will have benefited from the program. Similarly, half of the schools in Louga and Diourbel will have benefited from the program for CI, and none yet for CP. Questions about whether or not to sample from CP will need to be answered for the midline. This will have implications for interpretation because the sampling frame at baseline was done across all schools in all six regions.

One of the important student characteristics associated with positive student outcomes on Indicator 4, daily use of books in schools, is a pedagogical practice that can be stressed through professional development.

Concerning the instrument, most of the subtasks used for the midline EGRA should be those developed at baseline, in order to allow for a direct comparison of student performance. The listening comprehension task in Pulaar is an exception, and the 2018 version should be used, as explained above. In order to balance the benefit of more information and reporting to various indicators with the burden posed on students and the data collection process, a discussion should be held on the pertinence of adding a subtask in national language aimed to measure CI reading (the connected text from the EGRA CI study) and a French reading fluency subtask at CP. Furthermore, there should be an analysis of the instrument used at EGRA baseline (and that should be used again at midline and endline for comparison purposes) in relation to the level of text that will be taught with the Grade 2 materials now developed with Lecture Pour Tous support and with the preliminary benchmarks in mind, in case there are any important differences. Changing the text at midline while maintaining a measure to compare at baseline would require an additional pilot study before the midline data collection during which students would read both texts in order to equate the two measures, and with reading skills of the sample at the beginning level, it may be challenge to get sufficient data for equating.

Reporting on Indicators 1 and 4 of the AMELP require different student populations and subtasks. The current data collection plan is designed to report on Indicator 1 in program Years 3 and 5 (using midline and endline data) and Indicator 4 in Years 2 and 4 (using EGRA CI data). If the

program decides to report Indicator 4 in Years 3 and 5, an additional subtask would need to be included for CI students in those years (as mentioned above as an option).

Finally, considerations should be made as to how many contextual interview questions to include as essential questions were selected from baseline findings and included in the EGRA CI survey instruments. A similar selection of items could be included in the midline, which would shorten the test administration without lessening the robustness of the findings. Alternatively, the program may decide to use this data collection to collect additional information from students, teachers and school directors to help answer specific questions of interest for the program.

Annexes

Table A1. Wolof student accuracy scores by subtask and gender

Subtask	Male		Female	
	Mean	95% CI	Mean	95% CI
Listening Comprehension	71.0	66.7 ; 75.4	71.6	67.7 ; 75.4
Letter Sounds	64.0	59.5 ; 68.6	66.0	60.9 ; 71.0
Familiar Words	30.1	25.6; 34.6	27.4	22.6; 32.1
Connected Text Passage	29.4	25.0; 33.7	27.1	22.1; 32.2
Reading Comprehension	17.8	14.3; 21.2	15.8	12.2; 19.4

Table A2. Wolof students' accuracy scores by subtask and IA

	Fatick	Kaffrine	Kaolack	Matam
Subtask	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
Listening Comprehension	70.3 (65.3 ; 75.3)	71.9 (67.9 ; 76.0)	71.8 (66.1 ; 77.6)	50.3 (42.7 ; 58.0)
Letter Sounds	78.1 (72.7 ; 83.6)	57.3 (50.6 ; 64.1)	65.2 (57.3 ; 73.1)	43.9 (23.5 ; 64.4)
Familiar Words	43.3 (34.8 ; 51.9)	19.1 (14.7 ; 23.5)	29.6 (22.8 ; 36.3)	6.5 (0.0 ; 13.6)
Connected Text Passage	41.2 (32.7 ; 49.7)	19.4 (15.1 ; 23.7)	29.2 (22.5 ; 35.9)	10.5 (0.0 ; 22.2)

Reading Comprehension	25.1 (17.6 ; 32.7)	9.6 (6.9 ; 12.2)	18.6 (13.4 ; 23.6)	4.9 (0.0 ; 11.8)

Table A3. Pulaar student accuracy scores by subtask and gender

Subtask	Male		Female	
	Mean	95% CI	Mean	95% CI
Listening Comprehension	76.7	71.9 ; 81.4	77.2	72.7 ; 81.7
Letter Sounds	65.2	59.6 ; 70.9	60.4	53.2 ; 67.5
Familiar Words	21.3	15.9 ; 26.7	18.1	13.0 ; 23. 1
Connected Text Passage	11.4	0.6 ; 17.1	7.7	4.4 ; 10.8
Reading Comprehension	8.6	0.4 ; 13.3	6.4	3.3 ; 9.5

Table A4. Pulaar students' accuracy scores by subtask and IA

	Fatick	Kaffrine	Kaolack	Matam
Subtask	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
Listening Comprehension	50.9 (40.6 ; 61.2)	65.8 (60.1 ; 71.4)	64.9 (58.0 ; 71.8)	79.5 (75.1 ; 83.9)
Letter Sounds	70.9 (40.6 ; 60.8)	50.2 (39.9 ; 60.6)	61.4 (55.3 ; 67.6)	63.3 (56.8 ; 69.8)
Familiar Words	31.7	22.2	27.1	18.4

	(22.5 ; 40.9)	(13.6 ; 30.7)	(20.7 ; 33.6)	(13.5 ; 23.3)
Connected Text Passage	23.3 (14.3 ; 32.3)	12.3 (5.4 ; 19.3)	17.4 (11.1 ; 23.7)	8.1 (4.4 ; 11.9)
Reading Comprehension	13.3 (6.2 ; 20.4)	8.8 (4.0 ; 13.6)	11.3 (6.9 ; 15.6)	6.9 (3.5 ; 10.2)

Table A5. Seereer student accuracy scores by subtask and gender

Subtask	Male		Female	
	Mean	95% CI	Mean	95% CI
Listening Comprehension	66.3	60.7 ; 71.8	68.0	63.6 ; 72.4
Letter Sounds	53.7	47.9 ; 59.4	61.3	55.5 ; 67.2
Familiar Words	16.6	12.1 ; 20.9	25.6	19.2 ; 31.9
Connected Text Passage	22.8	17.1 ; 28.4	34.4	26.9 ; 41.9
Reading Comprehension	17.6	12.4 ; 22.8	29.4	21.2 ; 37.7

Table A6. Seereer students' accuracy scores by subtask and region

	Fatick	Kaffrine	Kaolack	Matam
Subtask	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
Listening Comprehension	68.2 (64.4 ; 71.9)	72.4 (62.7 ; 82.0)	63.6 (52.7 ; 74.6)	-
Letter Sounds	54.9	47.8	66.2	-

	(48.3 ; 61.5)	(25.7 ; 69.9)	(61.0 ; 71.4)	
Familiar Words	19.1 (13.3 ; 24.8)	25.8 (6.7 ; 45.0)	27.2 (20.6 ; 33.7)	-
Connected Text Passage	25.9 (18.9 ; 33.0)	29.9 (8.7 ; 51.2)	36.7 (30.1 ; 43.2)	-
Reading Comprehension	21.3 (14.9 ; 27.6)	23.9 (3.5 ; 44.3)	30.6 (19.9 ; 41.3)	-

Table A7. Comparison of Connected Texts in 2017 and 2018 Studies

	2017 EGRA Baseline (CI et CP)			2018 EGRA Light (CI)		
CONTENT	Wolof	Pulaar	Seereer	Wolof	Pulaar	Seereer
Type of text	narrative	narrative	narrative	narrative	narrative	narrative
Story Summary	Aadama's older sister goes to visit grandmother, grandmother is sick, Aadama's older sister gets lost, she asks for directions, someone gives her directions.	It's the beginning of the school year and Sala gets up early. Sala wakes up her younger sibling Aaali to leave for school. Sala wears new clothes. On the way, Sala falls, her robe gets dirty. She gets up, brushes off the sand, and starts walking again. They arrive at school on time. They greet the teacher and sit down.	Siga wants to play with her friend, she asks her aunt, she says not to stay long, she goes to play. When she sees it's getting late, she hurries back home. Her aunt is happy.	Badu has a ball and he loans it to his brother. His brother takes to school and gives it back when done.	Aaali has a sheep. Aali washes her/his sheep. He/she gives him/her some grass. Aali's sheep is running around	Ami wants to tress her hair. She is going to Dakar tomorrow. Her father sent her. Her sister braids her hair. She looks nice.
Presence of which letters in curricular sequence	All letters/graphemes in the alphabet.	All letters/graphemes in the alphabet.	All letters/graphemes in the alphabet.	Letter/graphemes through Week 15 in Year 1 curriculum.	Letter/graphemes through Week 15 in Year 1 curriculum.	Letter/graphemes through Week 15 in Year 1 curriculum.
LENGTH of TEXT						
Total number of words	54	46	45	19	16	20
Total number of sentences	6	7	6	**4**	**4**	**5**
Average no. words per sentence	9	6.5	7.5	4.75	4	4

Table A8. Comparison of 2017 and 2018 subtasks

Criteria	EGRA CI et CP (Baseline 2017)			EGRA CI (2018)		
Letter Sounds	Wolof	Pulaar	Seereer	Wolof	Pulaar	Seereer
	All graphemes included. 100 items on 10X10 grid.	All graphemes included. 100 items on 10X10 grid.	All graphemes included. 100 items on 10X10 grid.	Graphemes through Week 15 of Year 1 LPT curriculum. 50 items on 10X5 grid.	Graphemes through Week 15 of Year 1 LPT curriculum. 50 items on 10X5 grid.	Graphemes through Week 15 of Year 1 LPT curriculum. 50 items on 10X5 grid.
Familiar Words						
	Developed frequency lists based on any materials Lecture Pour Tous could provide in January-February 2017. Participants selected 50 items for the stimulus. 50 items on 5X10 grid	Developed frequency lists based on any materials Lecture Pour Tous could provide in January-February 2017. Participants selected 50 items for the stimulus. 50 items on 5X10 grid	Developed frequency lists based on any materials Lecture Pour Tous could provide in January-February 2017. Participants selected 50 items for the stimulus. 50 items on 5X10 grid	Chose a selection from baseline subtask and cross-checking with materials developed for use by Lecture Pour Tous by January 2018 (Year 1 LPT curriculum). 25 items on 5X5 grid.	Chose a selection from baseline subtask and cross-checking with materials developed for use by Lecture Pour Tous by January 2018 (Year 1 LPT curriculum). 25 items on 5X5 grid.	Chose a selection from baseline subtask and cross-checking with materials developed for use by Lecture Pour Tous by January 2018 (Year 1 LPT curriculum). 25 items on 5X5 grid.
Listening Comprehension						
	Developed by participants in workshop in January-February 2017.	Developed by participants in workshop in January-February 2017.	Developed by participants in workshop in January-February 2017.	Developed by participants in workshop in January-February 2017.	Developed by participants in workshop in January-February 2018 with intent to eliminate regional vocabulary and be intelligible to Pulaar-speaking children across the regions in Senegal.	Developed by participants in workshop in January-February 2017.

Table A9. Preliminary Benchmarks

Composante		Niveau	Niveaux de performance				Seuil minimal
			Débutant	Émergent	Compétent	Performant	
Compréhension	Compréhension à l'oral/à l'audition	CI et CP	0 à 40%	41% à 69%	70% à 90%	91% et plus	70%
	Compréhension du texte écrit (compréhension en lecture)	CI et CP	0 à 20%	21% à 59%	60% à 80%	81% et plus	60%
Lecture, texte continu	Fluidité de lecture d'un texte continu	CI et CP	0 à 5 mclm	6 à 24 mclm	25 à 35 mclm	36 et + mclm	25 mclm
	Précision de lecture d'un texte continu	CP	0% à 30%	31% à 69%	70% à 80%	81% et +	70%
		CI	0% à 25%	26% à 49%	50% à 70%	71% et +	50%
Lecture des mots familiers	Fluidité de lecture des mots familiers	CI, CP	0 à 4 mclm	5 à 19 mclm	20 à 30 mclm	31 et + mclm	20 mlcm
	Précision de lecture des mots familiers	CP	0% à 25%	26% à 59%	60% à 75%	76% et +	60%
		CI	0% à 19%	20% à 49%	50% à 70%	71% et +	50%
Lecture des mots par décodage	Fluidité de lecture des mots inventés	CI, CP	0 à 3 mclm	4 à 14 mclm	15 à 25 mclm	26 et + mclm	15 mclm
	Précision de lecture des mots inventés	CP	0% à 20%	21% à 49%	50% à 65%	66% et +	50%
		CI	0% à 20%	21% à 49%	50% à 65%	66% et +	50%
Conscience alphabétique	Fluidité de lecture des lettres et des syllabes	CI, CP	0 à 25 lclm	26 à 49 lclm	50 à 70 lclm	71 et + lclm	50 lclm
	Précision de lecture des lettres et syllabes	CP	0% à 30%	31% à 69%	70% à 90%	91% et +	70%
		CI	0% à 30%	31% à 69%	70% à 90%	91% et +	70%